

Impacto de Erros de Canal na Percepção de Qualidade em Sequências de Videoconferência de Alta Definição

Cassius D. Estrada, Alexandre Ciancio, José F. L. de Oliveira, Felipe Ribeiro Lopes, Eduardo A. B. da Silva, Amir Said

Resumo—Este artigo apresenta um estudo sobre a percepção de qualidade em sistemas de videoconferência de alta definição na presença de erros de canal (perda de blocos). De forma a conduzir este estudo, foi gerado um banco contendo 31 sequências de alta definição (1080i) sem compressão com diferentes características, tais como fundos de complexidade distinta, quantidade de indivíduos, entre outras. 16 sequências representativas foram corrompidas com erros de canal simulando perda de blocos que variou entre 0 e 0,05% do número total de blocos da sequência, e posterior ocultamento de erros. A análise dos resultados obtidos sugere que a natureza localizada dos erros de canal (em contraste com artefatos gerados durante a codificação) pode causar distorções na percepção de qualidade por parte do usuário que não estão associadas somente com a taxa bruta de erro, mas também com a natureza do conteúdo.

Palavras-Chave—HDTV, videoconferência, erros de canal, avaliação de qualidade.

Abstract—In this work we assess the impact of channel errors on the perceived quality in high definition videoconference applications. A database containing 31 high definition sequences (1080i) without compression was generated. The set of generated sequences may present backgrounds of different complexities as well as other features, such as number of subjects, face occlusion, skin color, among others. 16 representative sequences were degraded with channel error simulating packet loss that ranged from 0 to 0.05% and post processed by error concealment techniques. Results suggest that the localized nature of channel errors (in contrast to encoding errors) may mask the perception of quality by the users in a way that are not associated to the error rate itself, but also to the nature of the video content.

Keywords—HDTV, videoconference, channel errors, packet loss, quality assessment.

I. INTRODUÇÃO

Sistemas multimídia de alta qualidade têm se tornado cada vez mais presentes em nosso dia-a-dia, seja sob a forma de TV de Alta Definição, home-theaters, sistemas de telepresença ou mesmo telemedicina. Nestes cenários, atenção especial

deve ser dada a sistemas e aplicações que permitem interação humana em tempo real, tais como videoconferências. Sistemas comerciais de alto nível tais como o Halo, da HP, e o TelePresence, da Cisco, tentam prover ao usuário final uma experiência próxima à da presença física. Para atingir elevados níveis de satisfação e oferecer uma experiência verdadeiramente realista, tais sistemas devem enfrentar os compromissos entre os recursos disponíveis, as limitações do canal de transmissão, e a qualidade oferecida. Uma característica particular dos sistemas de telepresença de alta qualidade é que tais sistemas apresentam, de forma geral, poucos ou nenhum artefato de compressão (devido às altas taxas de bit utilizadas), estando sujeitos apenas a erros de transmissão, ou seja, à perda de blocos devido a falhas no canal.

A avaliação da qualidade de um sinal de vídeo contendo distorções pode ser feita tanto de forma objetiva (através de métricas de avaliação) quanto subjetiva (baseada na opinião de usuários). Diversas métricas para avaliação objetiva da qualidade de vídeo foram descritas na literatura e encontram-se disponíveis atualmente [1]–[4]. Entretanto, tais métricas foram, em grande parte, desenvolvidas visando avaliar distorções introduzidas exclusivamente durante o processo de codificação. Assim, uma vez que a natureza dos artefatos causados por erros de transmissão (em geral bem localizados no tempo e limitados a uma região específica do vídeo) é consideravelmente distinta dos artefatos de compressão (que afetam o vídeo como um todo), tais métricas são consideradas inadequadas para avaliação da qualidade percebida em sistemas com erro de canal. Estudos mostram que, na presença de erros de canal, tais métricas apresentam uma tendência de super ou subestimar o impacto percebido na qualidade [5]. Apesar de haver algumas métricas adaptadas à avaliação do efeito de erros de canal (a exemplo de [6]), seu desempenho em sistemas orientados para aplicações de videoconferência de alta definição é pouco conhecido.

A avaliação subjetiva consiste na realização de experimentos subjetivos (psico-físicos) com participantes (observadores) não-especialistas, com o objetivo de obter estimativas para a qualidade de vídeo de um conjunto de sequências. Estes testes subjetivos são considerados a forma mais precisa de aferir a

Cassius D. Estrada, Alexandre Ciancio, José F. L. de Oliveira, Felipe Ribeiro Lopes, Eduardo A. B. da Silva, PEE/COPPE e DEL/Poli, Universidade Federal do Rio de Janeiro, C. P. 68504 - Rio de Janeiro - RJ, 21941-972, Email: [cassius,ciancio,jleite,felipe.ribeiro,eduardo]@lps.ufrrj.br. Amir Said, Multimedia Communications and Networking Lab, HP Labs, 1501 Page Mill Road, MS 1811, USA, Email: amir_said@hp.com.

qualidade percebida de uma sequência de vídeo. Alguns estudos baseados em testes subjetivos sobre o impacto de erros de canal na percepção de qualidade já foram realizados. Em [7], o impacto da perda de pacotes em sistemas codificados com o padrão H.264 foi avaliado. Entretanto, tal estudo não considerou aplicações de alta definição e telepresença, concentrando-se em aplicações móveis. Já em [8], testes subjetivos para a avaliação de qualidade em sistemas HDTV foram conduzidos. Porém, neste caso, o estudo proposto avaliou somente sinais afetados por artefatos de compressão.

Neste artigo, avaliamos o impacto de erros de canal na percepção de qualidade em sequências de videoconferência de alta definição. Devido à carência de métricas objetivas reconhecidamente efetivas para a avaliação de qualidade em tal cenário, optamos pela condução de testes subjetivos para a avaliação da percepção de qualidade pelos usuários. Além disso, não há também na literatura um número adequado de sequências de vídeo disponível com as características desejadas. Assim, para a realização do estudo proposto, um banco de sequências de videoconferência com alta qualidade foi criado a partir de gravações de cenas em alta definição (1080i) com muito pouca compressão (perceptualmente sem perdas) realizadas em um estúdio da emissora TV Futura utilizando equipamentos profissionais de alta qualidade. A organização e condução dos testes subjetivos seguiu as normas estabelecidas pela ITU [9], [10]. Os resultados sugerem que em aplicações voltadas para telepresença, para uma dada taxa de erro de canal, a percepção de qualidade do usuário pode ser influenciada pela natureza do conteúdo presente no sinal de vídeo.

Este artigo está organizado do seguinte modo: na Seção II o banco de sequências gerado é descrito; na Seção III os procedimentos adotados na condução dos testes subjetivos, os casos de teste considerados e a metodologia de processamento dos dados obtidos são apresentados; na Seção IV os resultados obtidos são apresentados e comentados; a Seção V conclui o artigo.

II. SEQUÊNCIAS DE TESTE

31 sequências de videoconferência de alta qualidade e alta definição (1080i) foram geradas sem compressão utilizando equipamento profissional (câmera Sony PDW-700) em um estúdio da TV Futura no Rio de Janeiro. Tais sequências possuem três tipos de fundo (liso, simples e complexo) que variam em conteúdo de frequência, e apresentam características tais como movimento de mãos, rotação de faces, oclusão parcial de face, padrões de alta frequência (camisas listradas), indivíduos do sexo masculino e feminino, de diferentes cores de pele, e presença de óculos e barba. Quadros representativos de algumas sequências podem ser vistos na Fig. 1. Acreditamos que tal banco seja um recurso útil para a comunidade que trabalha com avaliação de qualidade de vídeo. As sequências

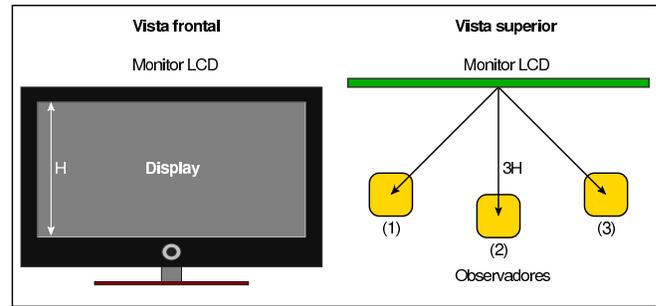


Fig. 2. Disposição dos observadores e do *display*.

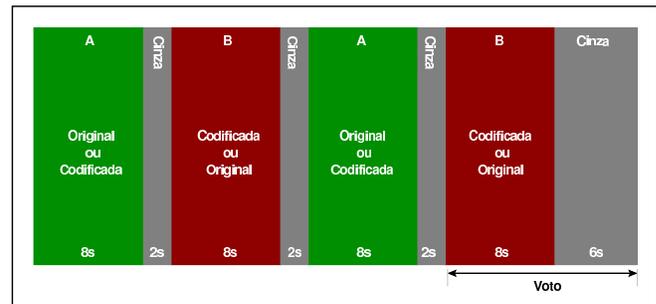


Fig. 3. Diagrama representativo da apresentação das sequências de teste no método DSCQS.

têm duração de 10 segundos e podem ser obtidas gratuitamente para uso não comercial em [11].

III. TESTES SUBJETIVOS

Neste trabalho, utilizamos os procedimentos normatizados para avaliação subjetiva estabelecidos pela ITU [9], [10]. Estes documentos fornecem especificações para as condições de visualização dos vídeos (que incluem a distância do observador para o monitor ou *display* e o controle da luz ambiente), critérios para seleção dos observadores e do material de teste, metodologia empregada nos experimentos e métodos para análise estatística dos dados coletados. A disposição dos observadores em relação ao monitor utilizou a configuração apresentada na Figura 2. Até três pessoas participaram de cada sessão de teste simultaneamente, sendo colocadas lado a lado, em frente ao monitor. Durante os testes, a sala tinha as luzes apagadas e janelas cobertas por cartolina preta. Este formato é semelhante ao utilizado pelo VQEG nos experimentos da Fase I [12].

Quinze observadores (não-especialistas na área de processamento de imagens ou vídeo) foram recrutados no campus da Universidade Federal do Rio de Janeiro, de forma a facilitar a geração de resultados estatisticamente confiáveis. O processo de avaliação seguiu a metodologia DSCQS – *Double Stimulus Continuous Quality Scale*. No método DSCQS, os observadores assistem a pares de sequências que consistem da sequência de teste (degradada) e da original correspondente. As sequências são apresentadas duas vezes, alternadamente



(a) Sequência com fundo liso.



(b) Sequência com fundo simples.



(c) Sequência com fundo simples e dois indivíduos.



(d) Sequência com fundo complexo.

Fig. 1. Quadros representativos de algumas sequências usadas nos experimentos.

(original e teste), conforme apresentado na Figura 3. No DSCQS, os observadores não são informados sobre quais as sequências são as originais e quais são as de teste. A cada sessão, as posições das sequências originais e das sequências de teste são modificadas aleatoriamente. As sequências são apresentadas em toda a tela da TV. O tom de cinza mostrado entre sequências tem luminância igual a 128. A varredura das linhas do vídeo está em sincronismo com a varredura do monitor. No período marcado como “voto” na Figura 3, o observador pode registrar o seu julgamento, ou seja, a sua nota para a qualidade geral de ambas as sequências de vídeo. A escala apresentada na Figura 4 (adaptação da Figura 6 da Seção 5.4 de [9]) é a escala de qualidade utilizada pelo DSCQS. As notas foram registradas em folhas de papel. O usuário atribui uma nota contínua (variando de ruim a excelente) a cada uma das duas sequências de forma individual, o que é feito inserindo uma marca em cada uma das escalas verticais (A e B).

A. Casos de Teste

Para a realização dos testes, 16 sequências representativas (SRCs – Sources) foram selecionadas do banco contendo as 31 sequências geradas. Estas sequências incluem 6 com fundo plano, 6 com fundo simples e 4 com fundo complexo, e correspondem às sequências de número 1, 3, 4, 5, 6, 7, 11, 14, 16, 17, 18, 24, 25, 28, 30, e 31 do banco de sequências disponibilizado em [11]. A simulação da quantidade de erros

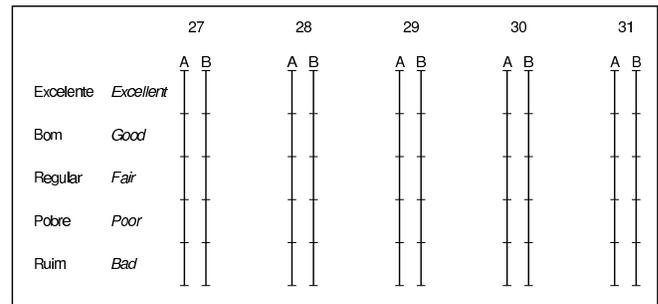


Fig. 4. Escala de níveis de qualidade utilizada pelo DSCQS.

de canal assumiu 100 valores distintos, representando 100 HRCs (*Hypothetical Reference Circuits*), ou casos de teste. Assim, cada HRC simulou uma porcentagem de perda de blocos variando lineamente entre 0 e 0,05% (em torno de 1220 blocos perdidos por sequência), resultando em um total de 1600 sequências degradadas. Os blocos perdidos foram distúrbios aleatoriamente em cada frame, e reconstruídos utilizando a estratégia de ocultamento de erros porposta segundo o padrão H.264 de codificação, que gera a informação de um bloco perdido através da interpolação dos dados de blocos vizinhos [13].

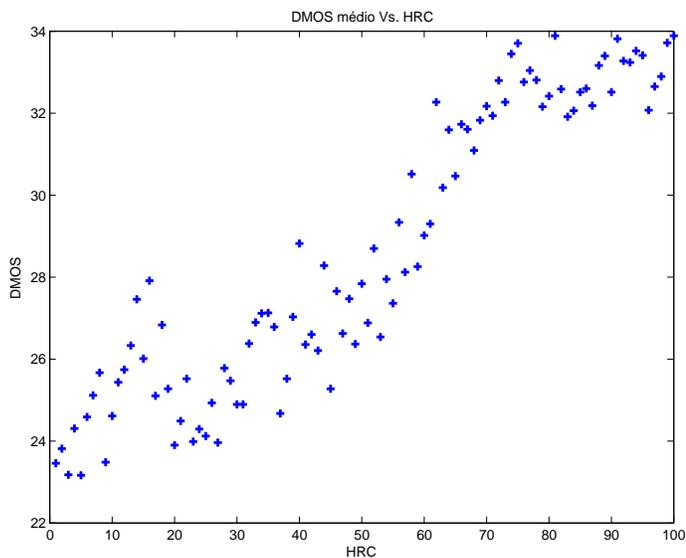


Fig. 5. DMOS médio, considerando todas as sequências, em função da taxa de erro.

B. Processamento dos Dados

As notas obtidas nos testes subjetivos foram processadas com base em [1]. A *nota diferencial* consiste na diferença entre a nota atribuída à sequência de referência e a nota atribuída à sequência de teste. A escala de notas dada através da folha de respostas variava entre 0 e 100. As notas foram normalizadas para cada usuário considerando todas as notas diferenciais dadas por este usuário em todos os testes. Após a normalização, notas fora do intervalo de confiança de 90% foram desconsideradas com base na análise de variância (ANOVA) dos dados. Por fim, a *nota diferencial média* (DMOS – *Differential Mean Opinion Score*) considerando as notas válidas foi calculada para cada caso de teste, gerando a nota (DMOS) final. Cabe ressaltar que, como o DMOS se baseia na diferença média entre as notas dadas à sequência de referência e à de teste, valores pequenos estão associados à boa qualidade, enquanto valores maiores indicam qualidade ruim. Os resultados obtidos em forma tabular também estão disponíveis em [11].

IV. RESULTADOS OBTIDOS

A Figura 5 mostra as notas (médias entre todas as sequências) obtidas para as diferentes taxas de perda de blocos (HRCs). Observa-se que, conforme o esperado, sequências com maior degradação tenderam a obter DMOSs maiores (pior qualidade). As correlações obtidas entre as notas atribuídas pelos usuários e as taxas de erro foram de 0,93 (correlação de Pearson) e 0,92 (correlação de Spearman).

Uma análise dos resultados por classe de fundo pode ser feita através do gráfico apresentado na Figura 6. Esta figura mostra a média de notas atribuídas a cada sequência considerando todos os casos de teste (HRCs). Percebe-se uma clara distinção de notas médias para as sequências com fundo

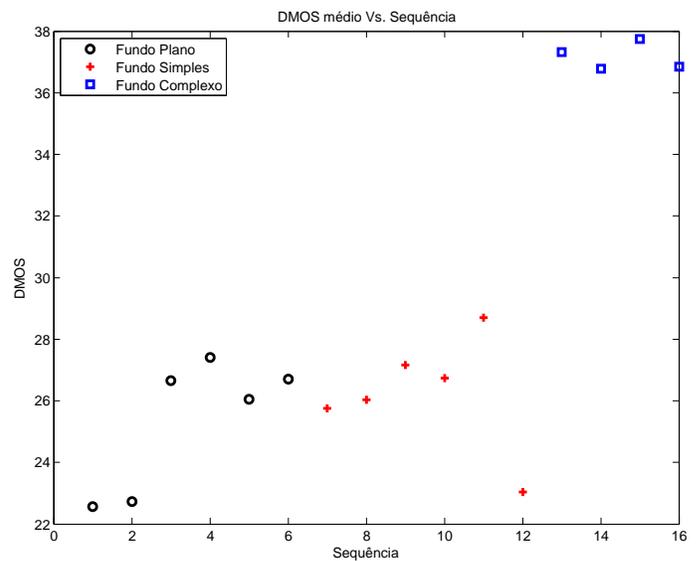


Fig. 6. DMOS médio para cada sequência de teste, considerando todas as taxas de erro testadas.

complexo (13 – 16). Este comportamento pode ser atribuído à natureza localizada dos erros de canal, que pode fazer com que a ocorrência de um erro fora da região de face, mas em uma região do fundo que contenha detalhes (alta frequência) desvie a atenção do usuário para aquela região, mascarando a real percepção de qualidade. Soma-se a isso o fato de que as sequências utilizadas nos testes não apresentavam áudio. Acredita-se que a ausência de vozes vindo das faces pode facilitar que a atenção do usuário seja desviada para regiões fora da face onde ocorram erros. Uma vez que a perda de blocos em sequências com fundo plano e simples não tende a desviar tanto a atenção dos usuários para essas regiões, tais sequências recebem, em geral, notas de qualidade melhores (DMOS médio menor). Desta forma, é interessante observar que a presença de conteúdo de alta frequência é um fator determinante na percepção de qualidade na presença de erros de canal pois, nestes casos, a atenção do usuário é desviada para as regiões de erro, alterando sua percepção de qualidade.

Na Figura 6 percebe-se também que as sequências 1, 2 e 12 (sequências de número 1, 3 e 24 em [11]) obtiveram valores mais baixos de DMOS (melhor qualidade) do que outras sequências com mesmo tipo de fundo. Tal efeito pode também ser explicado pela natureza de tais sequências. Nas sequências 1 e 2, a iluminação artificial mais intensa (em comparação às outras sequências de fundo plano) fez com que erros ocorridos na região do fundo fossem percebidos ainda com menor intensidade, melhorando a percepção de qualidade. Na sequência 12, um indivíduo manipulava blocos e outros objetos. Acredita-se que isto tenha feito com que os avaliadores concentrassem sua atenção nos objetos em movimento ao invés dos erros que ocorrem em outras regiões do vídeo.

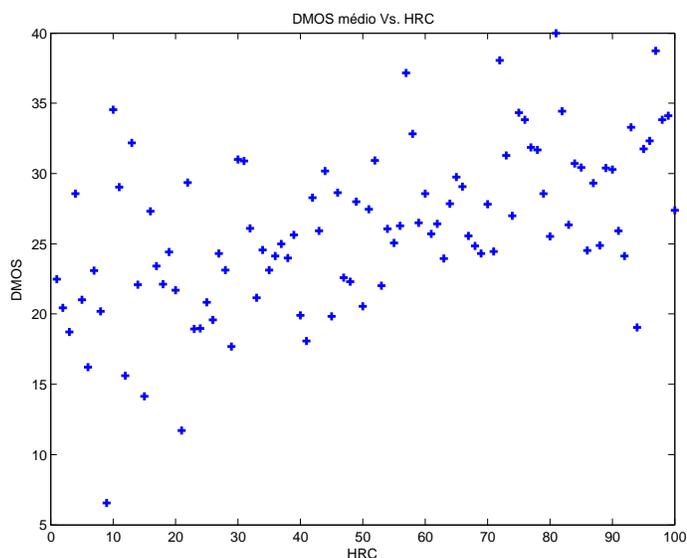


Fig. 7. DMOS médio para a sequência 5, em função da taxa de erro.

Outra observação interessante é que os gráficos de DMOS para as sequências individuais mostram uma correlação muito menor com a quantidade de degradação. A média dos coeficientes de correlação de Pearson para todas as sequências foi de 0,50, bem menor do que o obtido para a média de todas as sequências. Isto está ilustrado de Figura 7, onde é mostrado um gráfico dos valores de DMOS da sequência 5 por HRC, que obteve um coeficiente de correlação de Pearson de 0,55. Isto mostra que o desenvolvimento de uma métrica automática para avaliação de qualidade de videoconferências de alta definição na presença de erros de canal é uma tarefa longe de ser trivial, pois os fatores que afetam a qualidade são extremamente variáveis de uma sequência para outra. Entretanto, é importante ressaltar que o fato do DMOS médio entre as sequências (Figura 5) apresentar uma correlação razoavelmente alta com a quantidade de degradação sugere que a obtenção de uma métrica destas pode ser um objetivo atingível, e é um interessante tópico de pesquisa.

V. CONCLUSÕES

Neste artigo estudamos o impacto de erros de canal na percepção de qualidade em aplicações de videoconferência de alta definição. De forma a possibilitar o estudo, um banco de sequências de alta qualidade foi gerado e disponibilizado para a comunidade científica em [11]. A realização dos experimentos propostos consistiu na condução de testes subjetivos nos quais 15 usuários avaliaram 1600 sequências apresentando taxas de perda de bloco variando entre 0 e 0,05%. Estes resultados também estão disponíveis em [11]. A análise dos resultados obtidos sugere que a natureza localizada dos erros de canal (em contraste com artefatos gerados durante a codificação) pode causar distorções na percepção de qualidade por parte do usuário que não estão associadas somente com a

taxa bruta de erro, mas também com a natureza do conteúdo. Acreditamos que o bom entendimento dos mecanismos que influenciam a percepção de qualidade do usuário é de grande importância para o projeto e dimensionamento dos recursos associados à sistemas de vídeo, em especial aplicações voltadas à telepresença, e que vale a pena se investigar métricas automáticas de avaliação desta qualidade.

REFERÊNCIAS

- [1] VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment, Phase II," Tech. Rep., August 2003.
- [2] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, pp. 600–612, April 2004.
- [3] Z. Wang, H. R. Sheikh, and A. C. Bovik, "Objective video quality assessment," in *The Handbook of Video Databases: Design and Applications* (B. Furht and O. Marqure, eds.), pp. 1041–1078, Sept 2003, CRC Press.
- [4] S. Winkler, *Digital Video Quality: Vision Models and Metrics*, John Wiley & Sons, 2005.
- [5] U. Reiter, J. Korhonen, and J. You, "Comparing apples and oranges: assessment of the relative video quality in the presence of different types of distortions," *EURASIP Journal on Image and Video Processing*, vol. 8, 2011.
- [6] M. H. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," *IEEE Transactions on Broadcasting*, vol. 50, pp. 312–322, 2004.
- [7] M. Mu, R. Gostner, A. Mauthe, F. Garcia, and G. Tyson, "Visibility of individual packet loss on H.264 encoded video stream: A user study on the impact of packet loss on perceived video quality," in *Sixteenth Annual Multimedia Computing and Networking (MMCN'09)*, 2009.
- [8] J. Meenowa, D. S. Hands, R. Young, and D. Bayart, "Subjective assessment of HDTV content: comparison of quality across HDTV formats," in *Human Vision and Electronic Imaging*. 2010, vol. 7527 of *SPIE Proceedings*, SPIE.
- [9] International Telecommunication Union (ITU), "Recommendation ITU-R BT.500-11: Methodology for the subjective assessment of the quality of television pictures," 2002.
- [10] International Telecommunication Union (ITU), "Recommendation ITU-T P.910: Subjective video quality assessment methods for multimedia applications," 1999.
- [11] "High Quality HD Videoconference Database," <http://www.lps.ufrj.br/~tvdigital/futurasequences/>.
- [12] VQEG, "VQEG Phase I Subjective Test Plan," 1999.
- [13] J Mochnac and S. Marchevsky, "Error concealment scheme implemented in H.264/AVC," *50th International Symposium ELMAR*, 2008.